

Asynchronous Orthogonal Differential Modulation for Multiple Access Channels

Sina Poorkasmaei, *Student Member, IEEE*, and Hamid Jafarkhani, *Fellow, IEEE*

Abstract

We propose a differential encoding scheme and several differential decoding schemes for asynchronous multi-user MIMO systems based on orthogonal space-time block codes (OSTBCs) where neither the transmitters nor the receiver has knowledge of the channel. First, we derive novel low complexity differential decoders by performing interference cancelation in time and employing different decoding methods. The decoding complexity of these schemes grows linearly with the number of users. We then present additional differential decoding schemes that perform significantly better than our low complexity decoders and outperform the existing synchronous differential schemes, but require higher decoding complexity compared to our low complexity decoders. The proposed schemes work for any square OSTBC, any number of users, and any number of receive antennas. Furthermore, we analyze the diversity of the proposed schemes and prove that they achieve full diversity. Simulation results show that our differential schemes provide good performance. To the best of our knowledge, the proposed differential detection schemes are the first differential schemes for asynchronous multi-user systems.

Index Terms

Differential detection, multi-user detection, interference suppression, synchronization, space-time block coding.

The authors are with the Center for Pervasive Communications and Computing, University of California, Irvine, CA 92697 USA (e-mail: {spoorkas, hamidj}@uci.edu). This work was supported in part by the NSF Award CCF-0963925.

I. INTRODUCTION

Various space-time modulation techniques to achieve transmit diversity have been proposed in the literature [1]. In most cases, it is assumed that the channel state information (CSI) is perfectly known at the receiver [2], [3]. This is a reasonable assumption when the channel changes slowly and can be estimated by transmitting known training symbols. However, this is not always possible, and there is a tradeoff between frame length and accuracy of the channel estimation [4]. Therefore, the effects of channel estimation error make it desirable to use schemes that avoid such an estimation.

Prior work has proposed many differential space-time coding schemes in which neither the transmitter nor the receiver knows the CSI. The first differential coding schemes based on orthogonal designs for multiple transmit antennas were proposed in [5] and [6] with about 3-dB loss in performance compared to the corresponding coherent detection. Other examples of differential modulation schemes using space-time block codes (STBCs) and linear decoding complexity were proposed in [7], [8] and [9]. A rate-one differential modulation scheme based on the quasi-orthogonal space-time block codes (QOSTBCs) [10] can be found in [11].

Multi-user detection schemes with simple coherent detection structures for multiple access channels (MACs) have garnered significant attention [12], [13], [14]. The main goal is to design a low complexity interference cancelation method for a MAC with J users using only J receive antennas. This is done for $N = 2$ transmit antennas in [12] and for $J = 2$ users in [13] using the properties of orthogonal space-time block codes (OSTBCs) [3]. To solve the problem for any number of users, any constellation and any number of transmit antennas, [14] presents a method utilizing QOSTBCs with a moderate increase in decoding complexity. Space-time/frequency code design criteria for fading MIMO MACs and a code construction for two users have been derived in [15].

Differential modulation schemes for two-user MAC systems have been proposed in [16]. These

schemes have a high decoding complexity. In [17], we proposed low complexity differential modulation schemes for two-user MIMO systems that achieve full transmit diversity. Moreover, we presented additional differential decoding schemes that provide full diversity, outperform the existing differential schemes, and work for any square OSTBC.

All the existing multi-user differential schemes assume the transmission of the data by the users to be perfectly synchronized in time. To the best of our knowledge, a differential modulation scheme for asynchronous multi-user systems does not exist in the literature. In this paper, we design differential detection schemes for asynchronous multi-user MIMO systems where neither the transmitters nor the receiver knows the channel. Our main results are as follows:

- 1) With a slow Rayleigh fading channel model for an asynchronous multi-user system, we present a differential encoding algorithm and derive novel low complexity differential decoders by performing interference cancelation in time and employing different decoding methods. The decoding complexity of these schemes grows linearly with the number of users.
- 2) We also present additional differential decoding schemes that perform significantly better than our low complexity decoders and outperform the existing synchronous differential schemes, but need higher decoding complexity compared to our low complexity decoders.
- 3) All the proposed decoders work for any square OSTBC, any number of users, and any number of receive antennas.
- 4) We analyze the diversity of our schemes and prove that they all achieve full diversity. Simulation results show that the proposed differential detection schemes provide good performance.

The rest of the paper is organized as follows. In Section II, we introduce the system model. In Section III, we present the differential encoding for our asynchronous differential modulation schemes. The differential decoding schemes are put forward in Section IV. We prove that our

schemes achieve full diversity in Section V. Simulation results are provided in Section VI, and Section VII concludes the paper.

Notation: We use boldface capital letters to denote matrices, boldface small letters to denote vectors, and super-scripts $(\cdot)^*$ and $(\cdot)^\dagger$ to denote conjugate and conjugate transpose, respectively. $\|\cdot\|_F$ indicates the Frobenius norm, and $E[\cdot]$ represents the expected value. Also, we use \mathbf{I}_n and $\mathbf{0}_n$ to denote the $n \times n$ identity and zero matrices, respectively.

II. SYSTEM MODEL

We consider a wireless communication system with J users each with N transmit antennas and one receiver with M receive antennas with a quasi-static flat Rayleigh fading channel. We define \mathbf{H}_j , $j = 1, \dots, J$, as $M \times N$ channel fading matrices whose (m, n) th elements $h_{j,mn}$ are the channel fading coefficients from transmit antenna n to receive antenna m for User j . The entries of \mathbf{H}_j , $j = 1, \dots, J$, are samples of independent zero-mean complex Gaussian random variables with a variance of 0.5 per real dimension.

In a practical set-up, the transmitters use pulse-shaping filters, and the receiver usually utilizes a matched filter to maximize the SNR. In such a scenario, the role of the sampling is to provide a set of sufficient statistics for the detection of the received signals. Consider the signal vector transmitted by the j -th transmitter

$$\mathbf{x}_j(t) = \sum_k \mathbf{s}_j(k) \psi(t - kT_s) \quad (1)$$

where $\mathbf{s}_j(\cdot)$ is the $N \times 1$ symbol vector, T_s is a symbol duration, and $\psi(\cdot)$ is the pulse-shaping filter with a non-zero duration of at most LT_s for some $L \in \mathbb{N}$ (i.e., $\psi(t) = 0$, $|t| > \frac{L}{2}T_s$). We assume the average transmit power of each user is 1. The $M \times 1$ received signal is

$$\mathbf{y}(t) = \sum_{j=1}^J \mathbf{H}_j \mathbf{x}_j(t - \tau_j) + \mathbf{n}(t) = \sum_{j=1}^J \mathbf{H}_j \sum_k \mathbf{s}_j(k) \psi(t - kT_s - \tau_j) + \mathbf{n}(t) \quad (2)$$

where $\mathbf{n}(t)$ is the $M \times 1$ complex white Gaussian noise vector, and the symbol vectors $\mathbf{s}_j(k)$ for

the j -th user are transmitted through the channel matrix \mathbf{H}_j and received with a relative delay of τ_j . We assume τ_j is fixed within a frame. Then, considering the transmission of a frame of D symbol vectors $\mathbf{s}_j(1), \dots, \mathbf{s}_j(D)$ and assuming $\mathbf{s}_j(k) = 0$ for $k \notin \{1, \dots, D\}$, the optimum maximum-likelihood receiver uses the log-likelihood cost function given by

$$\begin{aligned} \Lambda &= \int \left\| \mathbf{y}(t) - \sum_{j=1}^J \mathbf{H}_j \sum_{k=1}^D \mathbf{s}_j(k) \psi(t - kT_s - \tau_j) \right\|_F^2 dt \\ &= \int \|\mathbf{y}(t)\|_F^2 dt + \int \left\| \sum_{j=1}^J \mathbf{H}_j \sum_{k=1}^D \mathbf{s}_j(k) \psi(t - kT_s - \tau_j) \right\|_F^2 dt \\ &\quad - 2 \operatorname{Re} \left\{ \operatorname{Tr} \left[\sum_{j=1}^J \left(\sum_{k=1}^D \int \mathbf{y}(t) \psi^*(t - kT_s - \tau_j) dt \cdot \mathbf{s}_j^\dagger(k) \right) \cdot \mathbf{H}_j^\dagger \right] \right\}. \end{aligned} \quad (3)$$

Now, consider the RHS of the last equality in (3). The first integral depends only on $\mathbf{y}(t)$, which is the same for all possible information sequences, and thus can be ignored for ML decoding. Also, for all possible information sequences in coherent detection, the second integral can be calculated at the receiver, independent of the received signal, since all its quantities are known. Finally, in terms of the received signal, it is sufficient to know only the last integral in order to perform ML decoding. Therefore, the output of the matched filter can be sampled at different sampling times associated with different transmitters to construct $\mathbf{y}_i(k)$ as follows

$$\mathbf{y}_i(k) = \int_{(k-\frac{L}{2})T_s+\tau_i}^{(k+\frac{L}{2})T_s+\tau_i} \mathbf{y}(t) \psi^*(t - kT_s - \tau_i) dt, \quad i = 1, \dots, J, \quad k = 1, \dots, D. \quad (4)$$

Clearly, the operations in (4) do not destroy any information that is valuable in deciding which symbols were transmitted, and thus these samples constitute a set of sufficient statistics for detecting all symbols. To simplify the notation, we assume that $\tau_1 = 0$, $\tau_1 < \tau_2 < \dots < \tau_J < T_s$, and $\tau_{(i_1+i_2 \cdot J)} = \tau_{i_1} + i_2 \cdot T_s$ ($\forall i_1, i_2 \in \mathbb{Z}$). We can write each integral in (4) as the sum of multiple integrals on smaller intervals. Then, we can scale the resulting integrals for simplification in notation and construct a new set consisting of all these integrals to obtain another set of sufficient

statistics for detection of all symbols as

$$\begin{aligned} \mathbf{y}_i(d) &= \frac{T_s}{\tau_{(i+1)i}} \int_{(d-\frac{L}{2})T_s+\tau_i}^{(d-\frac{L}{2})T_s+\tau_{i+1}} \mathbf{y}(t) \psi^*(t - dT_s - \tau_i) dt & i = 1, \dots, J, \\ &= \sum_{j=1}^J \mathbf{H}_j \sum_{r=0}^L s_j(d-r) \alpha_{j,i}(r) + \mathbf{n}_i(d), & d = 1, \dots, D+L, \end{aligned} \quad (5)$$

where $\tau_{i_1 i_2} = \tau_{i_1} - \tau_{i_2}$, $\forall i_1, i_2$,

$$\begin{aligned} \mathbf{n}_i(d) &= \frac{T_s}{\tau_{(i+1)i}} \int_{(d-\frac{L}{2})T_s+\tau_i}^{(d-\frac{L}{2})T_s+\tau_{i+1}} \mathbf{n}(t) \psi^*(t - dT_s - \tau_i) dt, \\ \alpha_{j,i}(r) &= \frac{T_s}{\tau_{(i+1)i}} \int_{(d-\frac{L}{2})T_s+\tau_i}^{(d-\frac{L}{2})T_s+\tau_{i+1}} \psi(t - (d-r)T_s - \tau_j) \psi^*(t - dT_s - \tau_i) dt \\ &= \frac{T_s}{\tau_{(i+1)i}} \int_{-\frac{L}{2}T_s}^{\tau_{(i+1)i}-\frac{L}{2}T_s} \psi(t + rT_s - \tau_{ji}) \psi^*(t) dt. \end{aligned} \quad (6)$$

Note that the last element of the set, $\mathbf{y}_J(D+L)$, is not obtained by splitting and scaling the integrals in (4). However, we make the notation simpler by adding it to the set, and the result is still a set of sufficient statistics. Also, notice that $\alpha_{j,i}(r) = 0$ for $r \notin \{0, \dots, L\}$. Therefore, the index r in (5) and (6) ranges from 0 to L . Moreover, $\mathbf{n}_i(d)$, $\forall i, d$, are independent zero-mean complex Gaussian random vectors with covariance matrices $E[\mathbf{n}_i(d) \mathbf{n}_i^\dagger(d)] = \frac{(\text{SNR})^{-1} T_s \alpha_{i,i}(0)}{\tau_{(i+1)i}}$. \mathbf{I}_M where SNR is the ratio of the average transmit power to the noise power. Let

$$\mathbf{Y}(d) = (\mathbf{y}_1(d), \dots, \mathbf{y}_J(d)), \quad \mathbf{N}(d) = (\mathbf{n}_1(d), \dots, \mathbf{n}_J(d)), \quad \boldsymbol{\alpha}_j(r) = (\alpha_{j,1}(r), \dots, \alpha_{j,J}(r)). \quad (7)$$

Then, the received samples can be written in a matrix form as

$$\mathbf{Y} = \sum_{j=1}^J \mathbf{H}_j \mathbf{S}_j \mathbf{A}_j + \mathbf{N} \quad (8)$$

where $\mathbf{Y} = (\mathbf{Y}(1), \dots, \mathbf{Y}(D+L))$, $\mathbf{S}_j = (s_j(1), \dots, s_j(D))$, $\mathbf{N} = (\mathbf{N}(1), \dots, \mathbf{N}(D+L))$ are $M \times (D+L)J$, $N \times D$ and $M \times (D+L)J$ matrices, respectively, and \mathbf{A}_j is a $D \times (D+L)J$

matrix given by

$$\mathbf{A}_j = \begin{pmatrix} \alpha_j(0) & \alpha_j(1) & \cdots & \alpha_j(L) & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \alpha_j(0) & \alpha_j(1) & \cdots & \alpha_j(L) & 0 & \cdots & 0 & \cdots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & \cdots & 0 & \alpha_j(0) & \alpha_j(1) & \cdots & \alpha_j(L) & 0 \\ 0 & \cdots & 0 & \cdots & 0 & 0 & \alpha_j(0) & \alpha_j(1) & \cdots & \alpha_j(L) \end{pmatrix}. \quad (9)$$

For the sake of simplicity, in this paper we consider the case where $L = 1$ and the pulse-shaping filter is a rectangular pulse

$$\psi(t) = \begin{cases} 1/\sqrt{T_s} & , -T_s/2 \leq t < T_s/2 \\ 0 & , \text{otherwise} \end{cases}. \quad (10)$$

Then, it can be easily seen from (6) that

$$\alpha_{j,i}(0) = \begin{cases} 1 & , j \leq i \\ 0 & , \text{otherwise} \end{cases}, \quad \alpha_{j,i}(1) = \begin{cases} 1 & , j > i \\ 0 & , \text{otherwise} \end{cases}. \quad (11)$$

Therefore, in this case, using (7), (9) and (11), \mathbf{A}_j becomes

$$\mathbf{A}_j = \begin{pmatrix} \overbrace{0 \cdots 0}^{j-1 \text{ times}} & \overbrace{1 \cdots 1}^{J \text{ times}} & 0 \cdots 0 & \cdots & 0 \cdots 0 & 0 \cdots 0 \\ 0 \cdots 0 & 0 \cdots 0 & \overbrace{1 \cdots 1}^{J \text{ times}} & \cdots & 0 \cdots 0 & 0 \cdots 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 \cdots 0 & 0 \cdots 0 & 0 \cdots 0 & \cdots & \overbrace{1 \cdots 1}^{J \text{ times}} & \overbrace{0 \cdots 0}^{J-j+1 \text{ times}} \end{pmatrix}, \quad (12)$$

and $\mathbf{n}_i(d)$, $\forall i, d$, become independent zero-mean complex Gaussian random vectors with covariance matrices $E[\mathbf{n}_i(d)\mathbf{n}_i^\dagger(d)] = \frac{(\text{SNR})^{-1}T_s}{\tau_{(i+1)i}} \cdot \mathbf{I}_M$.

In what follows, we consider the received signals in size TJ blocks of $(\mathbf{y}_1(Tl+1), \cdots, \mathbf{y}_J(Tl+1), \cdots, \mathbf{y}_1(Tl+T), \cdots, \mathbf{y}_J(Tl+T))$, for $l = 0, 1, \cdots$, and with a small abuse of the notation,

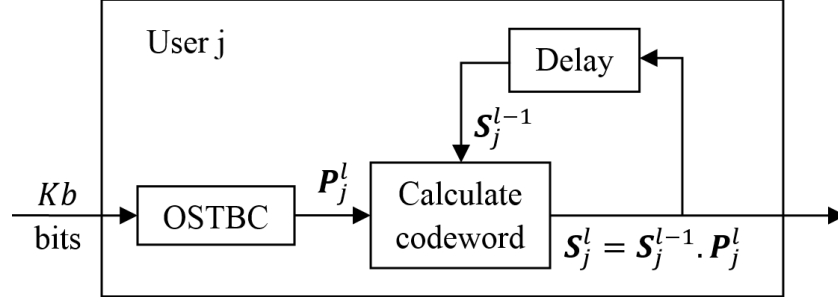


Fig. 1. Block diagram of differential encoder.

we denote them as $(\mathbf{y}_{1,1}^l, \dots, \mathbf{y}_{1,J}^l, \dots, \mathbf{y}_{T,1}^l, \dots, \mathbf{y}_{T,J}^l)$. Similarly, we denote the noise terms $(\mathbf{n}_1(Tl+1), \dots, \mathbf{n}_J(Tl+1), \dots, \mathbf{n}_1(Tl+T), \dots, \mathbf{n}_J(Tl+T))$ as $(\mathbf{n}_{1,1}^l, \dots, \mathbf{n}_{1,J}^l, \dots, \mathbf{n}_{T,1}^l, \dots, \mathbf{n}_{T,J}^l)$, for $l = 0, 1, \dots$. We define K as the number of data symbols transmitted during one block. The channels are assumed to be unknown at both the transmitters and the receiver.

III. DIFFERENTIAL ENCODING

In this section, we describe our differential encoding scheme for User $j = 1, \dots, J$. The block diagram of the differential encoder is the same as that of a synchronized system and is shown in Fig. 1. The main difference with the synchronous case [16], [17] is that different users do not need to employ different constellations. At a transmission rate of b bits/(s Hz), we use a same-length signal constellation with 2^b elements such as 2^b -PSK with an appropriate normalization to make the transmitted codewords unitary. Similar to the case of a single user, extension to other constellations is possible. For each block of Kb bits, User j selects K symbols and transmits them using an $N \times N$ OSTBC. This transmitted codeword also depends on the codeword and symbols transmitted in the previous block. We assume the input bits are the outputs of independent uniformly distributed random variables.

The encoding starts with the transmission of arbitrary $N \times N$ OSTBCs \mathbf{S}_j^0 and \mathbf{S}_j^1 . As in the case of a single user, we could transmit only one OSTBC instead of two and the system

would still work with minor changes. For block l , we use the Kb input bits to pick K symbols $p_{j,1}^l, \dots, p_{j,K}^l$ from the signal constellation and construct the corresponding square OSTBC, \mathbf{P}_j^l . Assuming that \mathbf{S}_j^{l-1} is the codeword of User j for the $(l-1)$ th block, we calculate \mathbf{S}_j^l by

$$\mathbf{S}_j^l = \mathbf{S}_j^{l-1} \cdot \mathbf{P}_j^l \quad (13)$$

and then transmit it at block l . Note that the generated codeword \mathbf{S}_j^l will be orthogonal as well.

With a small abuse of the notation, for data matrices $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4$, let us define

$$\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4) \triangleq \begin{pmatrix} \mathbf{I}_N & \mathbf{P}_1 & \mathbf{P}_1\mathbf{P}_2 \\ \mathbf{I}_N & \mathbf{P}_3 & \mathbf{P}_3\mathbf{P}_4 \end{pmatrix} \cdot \bar{\mathbf{A}} \quad (14)$$

where $\bar{\mathbf{A}}$ is a $3T \times 3T - 1$ matrix given later in (16). We choose the signal constellation such that for any possible data matrices $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4$ with $(\mathbf{P}_1, \mathbf{P}_2) \neq (\mathbf{P}_3, \mathbf{P}_4)$, the matrix $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$ has full row rank (i.e., $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$ is of rank $2N$). Later, in Section V, we show that under this condition all the proposed schemes provide full diversity. We also derive an equivalent condition, which can be easily verified using simple matrix operations.

IV. DIFFERENTIAL DECODING

In this section, we present differential decoding schemes for all users. First, we derive novel low complexity decoders by performing interference cancelation in time and employing different decoding methods. The decoding complexity of these decoders increases linearly with the number of users. We then present additional decoding schemes that perform significantly better compared to our low complexity decoders and outperform the existing synchronous differential schemes. All the proposed decoders work for any square OSTBC, any number of users, and any number of receive antennas. We assume that the channel is unchanged within three consecutive time blocks¹.

¹As will become clear later, the channel could be assumed to be unchanged within a shorter period of time, and our schemes would still work with minor changes.

A. Low Complexity Decoding Schemes

In this subsection, we introduce low complexity decoders for J users with N transmit antennas through several decoding methods. We illustrate the decoding process for User $j = 1, \dots, J$. In Method 0, we derive a low complexity decoder by canceling the interference of all users on User j and then performing maximum-likelihood decoding. Based on the decoder in Method 0, we then use Methods 1 and 2 presented in [17] to improve the performance. These methods use dynamic programming (DP) to efficiently decode the transmitted data signals. Finally, we present another decoding method (Method 3) to further reduce the decoding complexity while maintaining good performance.

Method 0: We use the following proposition to design our low complexity decoders:

Proposition 4.1: For any $l \geq 2$, the following relationship holds between the received signals and the transmitted signals of User $j = 1, \dots, J$

$$\bar{\mathbf{Y}}_j^l = \mathbf{H}_j \mathbf{S}_j^{l-2} \mathbf{U}_j^l \bar{\mathbf{A}} + \bar{\mathbf{N}}_j^l \quad (15)$$

where $\bar{\mathbf{A}}$ is a $3T \times 3T - 1$ matrix given by

$$\bar{\mathbf{A}} = \begin{pmatrix} -1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & \cdots & -1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & -1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \quad \begin{aligned} \bar{\mathbf{Y}}_j^l &= (\bar{\mathbf{y}}_2^{l-2}, \dots, \bar{\mathbf{y}}_T^{l-2}, \bar{\mathbf{y}}_1^{l-1}, \dots, \bar{\mathbf{y}}_T^{l-1}, \bar{\mathbf{y}}_1^l, \dots, \bar{\mathbf{y}}_T^l), \\ \bar{\mathbf{N}}_j^l &= (\bar{\mathbf{n}}_2^{l-2}, \dots, \bar{\mathbf{n}}_T^{l-2}, \bar{\mathbf{n}}_1^{l-1}, \dots, \bar{\mathbf{n}}_T^{l-1}, \bar{\mathbf{n}}_1^l, \dots, \bar{\mathbf{n}}_T^l), \\ \mathbf{U}_j^l &= (\mathbf{I}_N, \mathbf{P}_j^{l-1}, \mathbf{P}_j^{l-1} \mathbf{P}_j^l), \end{aligned} \quad (16)$$

with $\bar{\mathbf{y}}_t^l = \mathbf{y}_{t,j}^l - \mathbf{y}_{t,j-1}^l$, $\bar{\mathbf{n}}_t^l = \mathbf{n}_{t,j}^l - \mathbf{n}_{t,j-1}^l$ for $t = 1, \dots, T$ and $\forall l$ (assuming that $\mathbf{y}_{t,0}^l, \mathbf{n}_{t,0}^l$ denote $\mathbf{y}_{t-1,J}^l, \mathbf{n}_{t-1,J}^l$, respectively).

Proof: Using the input-output relationship in (8) and (12), we can write the input-output

relationship for a single time block $l > 0$ as

$$\left(\mathbf{y}_{1,1}^l, \dots, \mathbf{y}_{1,J}^l, \dots, \mathbf{y}_{T,1}^l, \dots, \mathbf{y}_{T,J}^l \right) = \sum_{i=1}^J \mathbf{H}_i \left(\mathbf{S}_i^{l-1}, \mathbf{S}_i^l \right) \begin{pmatrix} \mathbf{Z}_{i,1} \\ \mathbf{Z}_{i,0} \end{pmatrix} + \left(\mathbf{n}_{1,1}^l, \dots, \mathbf{n}_{1,J}^l, \dots, \mathbf{n}_{T,1}^l, \dots, \mathbf{n}_{T,J}^l \right) \quad (17)$$

where $\mathbf{Z}_{i,0}, \mathbf{Z}_{i,1}$, $i = 1, \dots, J$, are $T \times TJ$ matrices given by

$$\mathbf{Z}_{i,0} = \begin{pmatrix} \overbrace{0 \dots 0}^{i-1 \text{ times}} & \overbrace{1 \dots 1}^{J \text{ times}} & 0 \dots 0 & \dots & 0 \dots 0 \\ 0 \dots 0 & 0 \dots 0 & \overbrace{1 \dots 1}^{J \text{ times}} & \dots & 0 \dots 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 \dots 0 & 0 \dots 0 & 0 \dots 0 & \dots & \overbrace{1 \dots 1}^{J-i+1 \text{ times}} \end{pmatrix}, \quad \mathbf{Z}_{i,1} = \begin{pmatrix} 0 \dots 0 & 0 \dots 0 & 0 \dots 0 & \dots & 0 \dots 0 \\ 0 \dots 0 & 0 \dots 0 & 0 \dots 0 & \dots & 0 \dots 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \underbrace{1 \dots 1}_{i-1 \text{ times}} & \underbrace{0 \dots 0}_{TJ-i+1 \text{ times}} & 0 \dots 0 & \dots & 0 \dots 0 \end{pmatrix}. \quad (18)$$

Then, note that the interference of all users on User j can be canceled by subtracting $\mathbf{y}_{t,j-1}^l$ from $\mathbf{y}_{t,j}^l$ for $t = 1, \dots, T$ as follows

$$\left(\mathbf{y}_{1,j}^l - \mathbf{y}_{1,j-1}^l, \dots, \mathbf{y}_{T,j}^l - \mathbf{y}_{T,j-1}^l \right) = \mathbf{H}_j \left(\mathbf{S}_j^{l-1}, \mathbf{S}_j^l \right) \begin{pmatrix} \bar{\mathbf{Z}}_1 \\ \bar{\mathbf{Z}}_0 \end{pmatrix} + \left(\mathbf{n}_{1,j}^l - \mathbf{n}_{1,j-1}^l, \dots, \mathbf{n}_{T,j}^l - \mathbf{n}_{T,j-1}^l \right) \quad (19)$$

where $\bar{\mathbf{Z}}_0, \bar{\mathbf{Z}}_1$ are $T \times T$ matrices given by

$$\bar{\mathbf{Z}}_0 = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & \dots & -1 & 0 \\ 0 & 0 & 0 & \dots & 1 & -1 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}, \quad \bar{\mathbf{Z}}_1 = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ -1 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}. \quad (20)$$

Considering (19) for more consecutive time slots and using simple algebra, one may easily see that

$$\bar{\mathbf{Y}}_j^l = \mathbf{H}_j \cdot \left(\mathbf{S}_j^{l-2}, \mathbf{S}_j^{l-1}, \mathbf{S}_j^l \right) \cdot \bar{\mathbf{A}} + \bar{\mathbf{N}}_j^l = \mathbf{H}_j \mathbf{S}_j^{l-2} \mathbf{U}_j^l \bar{\mathbf{A}} + \bar{\mathbf{N}}_j^l. \quad (21)$$

■

Equation (15) is the main property for our low complexity differential decoding algorithm,

where the interference of all users on User j is completely canceled. Therefore, it can be utilized to decode the transmitted signals without interference. Notice that $\bar{\mathbf{Y}}_j^l$ starts from $\bar{\mathbf{y}}_2^{l-2}$ instead of $\bar{\mathbf{y}}_1^{l-2}$. We could consider using $\bar{\mathbf{y}}_1^{l-2}$ and other previously received signals to improve the performance of our scheme. However, that would cause additional inter-block interference from the previously transmitted signals of User j , which would then increase the decoding complexity. It is easy to see from (15) that when conditioned on $\mathbf{P}_j^{l-1}, \mathbf{P}_j^l$, the matrix $\bar{\mathbf{Y}}_j^l$ is Gaussian with conditional probability density function (pdf)

$$P(\bar{\mathbf{Y}}_j^l | \mathbf{P}_j^{l-1}, \mathbf{P}_j^l) \propto \frac{\exp\{-\text{Tr}[\bar{\mathbf{Y}}_j^l \cdot (\bar{\mathbf{V}}_j^l)^{-1} \cdot (\bar{\mathbf{Y}}_j^l)^\dagger]\}}{[\det(\bar{\mathbf{V}}_j^l)]^M} \quad (22)$$

where $\bar{\mathbf{V}}_j^l$ is the covariance matrix given by $\bar{\mathbf{V}}_j^l = (\mathbf{U}_j^l \bar{\mathbf{A}})^\dagger \cdot (\mathbf{U}_j^l \bar{\mathbf{A}}) + (\text{SNR})^{-1} T_s(\tau_{(j+1)j}^{-1} + \tau_{j(j-1)}^{-1}) \cdot \mathbf{I}_{3T-1}$. We are now prepared to present our first low complexity differential decoding scheme. One approach is to decode \mathbf{P}_j^{l-1} and \mathbf{P}_j^l jointly based on (22). We define the Inter-Time Interference Cancellation (ITIC) decoding using Method 0 as

$$\{\hat{\mathbf{P}}_j^{l-1}, \hat{\mathbf{P}}_j^l\} = \underset{\mathbf{P}_j^{l-1}, \mathbf{P}_j^l}{\text{argmin}} \Lambda_j^l(\mathbf{P}_j^{l-1}, \mathbf{P}_j^l) \quad (23)$$

where $\Lambda_j^l(\mathbf{P}_j^{l-1}, \mathbf{P}_j^l)$ is given by

$$\Lambda_j^l(\mathbf{P}_j^{l-1}, \mathbf{P}_j^l) = M \cdot \ln[\det(\bar{\mathbf{V}}_j^l)] + \text{Tr}[\bar{\mathbf{Y}}_j^l \cdot (\bar{\mathbf{V}}_j^l)^{-1} \cdot (\bar{\mathbf{Y}}_j^l)^\dagger]. \quad (24)$$

Notice that for $l = 2$ in \mathbf{U}_j^l , $\mathbf{P}_j^1 = (\mathbf{S}_j^0)^\dagger \mathbf{S}_j^1$ is the arbitrary data matrix at block 1 and is known at both the encoder and decoder. When using this scheme, information provided by (22) at time blocks other than l is ignored, and thus some performance is lost. To avoid such losses, we also propose additional decoding schemes using Methods 1 and 2 presented in [17] to efficiently decode the signals transmitted by the users. Note that we use the cost function of the ITIC decoder using Method 0 as described above, and thus the corresponding decoders using Methods 1 and 2 as presented in this paper are different from the decoders presented in

[17]. In what follows, we summarize the description of the ITIC decoders using Methods 1 and 2 based on the cost function of the ITIC decoder using Method 0. We refer the interested reader to [17] for the details on derivations.

Method 1 (Causal DP): In Method 1, we decode \mathbf{P}_j^l based on (22) for all blocks $\ell = 2, \dots, l$ together. We utilize DP to efficiently find the best possible data matrix that maximizes an approximation for the conditional pdf of $\bar{\mathbf{Y}}_j^0, \dots, \bar{\mathbf{Y}}_j^l$ given the data matrices $\mathbf{P}_j^2, \dots, \mathbf{P}_j^l$. Using (22) and ignoring the correlations of $\bar{\mathbf{Y}}_j^\ell$ at different blocks $\ell = 2, \dots, l$ given the data matrices, we consider the following:

$$f_1(\mathbf{P}_j^2, \dots, \mathbf{P}_j^l) \propto \prod_{\ell=2}^l \exp \left\{ -\Lambda_j^\ell(\mathbf{P}_j^{\ell-1}, \mathbf{P}_j^\ell) \right\} = \exp \left\{ -\sum_{\ell=2}^l \Lambda_j^\ell(\mathbf{P}_j^{\ell-1}, \mathbf{P}_j^\ell) \right\}. \quad (25)$$

In order to maximize the above function, we only need to minimize $\sum_{\ell=2}^l \Lambda_j^\ell(\mathbf{P}_j^{\ell-1}, \mathbf{P}_j^\ell)$. For any block $l \geq 2$, we define the ITIC decoding using Method 1 as

$$\hat{\mathbf{P}}_j^l = \underset{\mathbf{P}_j^l}{\operatorname{argmin}} \Phi_j^l(\mathbf{P}_j^l) \quad (26)$$

where $\Phi_j^l(\mathbf{P}_j^l)$ is defined as

$$\Phi_j^l(\mathbf{P}_j^l) \triangleq \begin{cases} \Lambda_j^2(\mathbf{P}_j^1, \mathbf{P}_j^2) & , l = 2 \\ \min_{\mathbf{P}_j^2, \dots, \mathbf{P}_j^{l-1}} \sum_{\ell=2}^l \Lambda_j^\ell(\mathbf{P}_j^{\ell-1}, \mathbf{P}_j^\ell) & , \text{otherwise} \end{cases}. \quad (27)$$

The optimization problem in (27) can be efficiently solved by utilizing DP. Using (27), it is easy to show that for $l > 2$, we have

$$\Phi_j^l(\mathbf{P}_j^l) = \min_{\mathbf{P}_j^{l-1}} \left\{ \Phi_j^{l-1}(\mathbf{P}_j^{l-1}) + \Lambda_j^l(\mathbf{P}_j^{l-1}, \mathbf{P}_j^l) \right\}. \quad (28)$$

As a result of storing the cost function of the previous block, $\Phi_j^{l-1}(\mathbf{P}_j^{l-1})$, we only need to perform an optimization over \mathbf{P}_j^{l-1} for each time block l . That is, in lieu of solving the optimization problem in (27) over all data matrices for the previous blocks, $\mathbf{P}_j^2, \dots, \mathbf{P}_j^{l-1}$, we

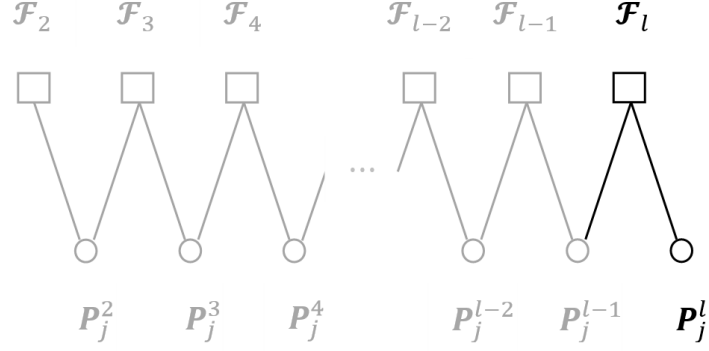


Fig. 2. Chain corresponding to the decoding of P_j^l .

can solve the optimization problem in (28) over the data matrix of only one block, P_j^{l-1} , as illustrated in Fig. 2. The optimization in (28) corresponds to the black path, while the optimization for the previous blocks corresponds to the gray path.

Method 2 (Non-Causal DP): In Method 2, we consider some non-overlapping windows of blocks and decode the transmitted symbols within each window together. Note that since the decoding of each block may depend on future blocks in the same window, this method will cause some additional delay. However, since more information is used, the performance will improve as well.

Using Method 2, in the m th stage of decoding, $m \geq 1$, we decode the data matrices at blocks $k_{m-1} + 1, \dots, k_m$ where $k_0 = 1$ and $k_0 < k_1 < k_2 < \dots$. We consider the following:

$$f_2(P_j^2, \dots, P_j^{k_m}) \propto \prod_{\ell=2}^{k_m} \exp \{ -\Lambda_j^\ell(P_j^{\ell-1}, P_j^\ell) \} = \exp \left\{ - \sum_{\ell=2}^{k_m} \Lambda_j^\ell(P_j^{\ell-1}, P_j^\ell) \right\}. \quad (29)$$

Then, in order to decode the data matrix for any block l ($k_{m-1} < l \leq k_m$), we use DP to find the best estimate of P_j^l that maximizes $f_2(P_j^2, \dots, P_j^{k_m})$ in (29). In order to maximize the above function, we only need to minimize $\sum_{\ell=2}^{k_m} \Lambda_j^\ell(P_j^{\ell-1}, P_j^\ell)$. Therefore, for any $m \geq 1$, we define

the m th stage of the ITIC decoding using Method 2 as

$$\left\{ \hat{\mathbf{P}}_j^{k_{m-1}+1}, \dots, \hat{\mathbf{P}}_j^{k_m} \right\} = \underset{\mathbf{P}_j^{k_{m-1}+1}, \dots, \mathbf{P}_j^{k_m}}{\operatorname{argmin}} \left\{ \min_{\mathbf{P}_j^2, \dots, \mathbf{P}_j^{k_{m-1}}} \sum_{\ell=2}^{k_m} \Lambda_j^\ell (\mathbf{P}_j^{\ell-1}, \mathbf{P}_j^\ell) \right\}. \quad (30)$$

To reduce the complexity of the exhaustive search in (30), we use dynamic programming as described below. Let us denote the minimizing arguments of $\sum_{\ell=2}^{k_m} \Lambda_j^\ell (\mathbf{P}_j^{\ell-1}, \mathbf{P}_j^\ell)$ by $\hat{\mathbf{P}}_j^2, \dots, \hat{\mathbf{P}}_j^{k_m}$.

If we know $\hat{\mathbf{P}}_j^{l+1}$ ($k_{m-1} < l \leq k_m - 1$), it can be easily shown that $\hat{\mathbf{P}}_j^l$ can be written as

$$\hat{\mathbf{P}}_j^l = \underset{\mathbf{P}_j^l}{\operatorname{argmin}} \left\{ \Phi_j^l (\mathbf{P}_j^l) + \Lambda_j^{l+1} (\mathbf{P}_j^l, \hat{\mathbf{P}}_j^{l+1}) \right\}. \quad (31)$$

Therefore, if we know $\hat{\mathbf{P}}_j^{l+1}$ and $\Phi_j^l (\mathbf{P}_j^l)$, we can compute $\hat{\mathbf{P}}_j^l$ using (31). This is the key element of our low complexity decoder using Method 2.

In the m th stage of decoding, similar to Method 1, we begin by employing (27) to compute and store $\Phi_j^\ell (\mathbf{P}_j^\ell)$, $\ell = k_{m-1}+1, \dots, k_m$, for any possible data matrix \mathbf{P}_j^ℓ using the stored values of $\Phi_j^\ell (\mathbf{P}_j^\ell)$ from the previous block. As in Method 1, once the signals for block ℓ are received, we can compute $\Phi_j^\ell (\mathbf{P}_j^\ell)$ with no additional delay. Note that $\hat{\mathbf{P}}_j^{k_m}$ is then exactly the same as in Method 1 because (25) and (29) (and therefore the resulting cost functions) are identical for decoding block $l = k_m$. Thus, at block k_m , we compute $\hat{\mathbf{P}}_j^{k_m} = \underset{\mathbf{P}_j^{k_m}}{\operatorname{argmin}} \Phi_j^{k_m} (\mathbf{P}_j^{k_m})$ as the best estimate of the data matrix $\mathbf{P}_j^{k_m}$, which then determines the decoded bits. We then move backwards, decoding the remaining matrices one at a time beginning from $\mathbf{P}_j^{k_m-1}$ and ending at $\mathbf{P}_j^{k_{m-1}+1}$ using (31), that is, utilizing the last decoded matrix and the stored values of $\Phi_j^\ell (\mathbf{P}_j^\ell)$, $\ell = k_{m-1} + 1, \dots, k_m - 1$. Finally, we supply the decoded bits for each time block.

Method 3 (Decision Feedback): An alternative approach to decoding \mathbf{P}_j^l at block l is to use the decoded matrix for \mathbf{P}_j^{l-1} at block $l - 1$ in (23). Therefore, we define the ITIC decoding using Method 3 as

$$\hat{\mathbf{P}}_j^l = \underset{\mathbf{P}_j^l}{\operatorname{argmin}} \Lambda_j^l (\hat{\mathbf{P}}_j^{l-1}, \mathbf{P}_j^l) \quad (32)$$

where $\hat{\mathbf{P}}_j^{l-1}$ is the decoded matrix for \mathbf{P}_j^{l-1} at block $l-1$. Notice that by using this approach, in order to decode \mathbf{P}_j^l we only need to solve an optimization over \mathbf{P}_j^l . Therefore, the decoding complexity is significantly reduced compared to the previous three decoding methods. However, the decoded signals for \mathbf{P}_j^{l-1} at block $l-1$ may be erroneous, which can lead to error propagation and thus performance degradation. We study the effect of error propagation in Section VI and show that it is not significant.

B. Optimal Multiple Partition Decoding Schemes

In this subsection, we present additional decoding schemes that achieve significantly higher coding gains compared to our low complexity schemes. In order to do this, we need the following proposition:

Proposition 4.2: For any $l \geq 2$, the following relationship holds

$$\tilde{\mathbf{Y}}^l = \sum_{j=1}^J \mathbf{H}_j \mathbf{S}_j^{l-2} \mathbf{U}_j^l \tilde{\mathbf{A}}_j + \tilde{\mathbf{N}}^l \quad (33)$$

where $\tilde{\mathbf{A}}_j$ is a $3T \times 3TJ - J + 1$ matrix given by

$$\tilde{\mathbf{A}}_j = \begin{pmatrix} \overbrace{1 \cdots 1}^{j \text{ times}} & 0 \cdots 0 & \cdots & 0 \cdots 0 \\ 0 \cdots 0 & \overbrace{1 \cdots 1}^{J \text{ times}} & \cdots & 0 \cdots 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 \cdots 0 & 0 \cdots 0 & \cdots & \overbrace{1 \cdots 1}^{J-j+1 \text{ times}} \end{pmatrix}, \quad (34)$$

$$\tilde{\mathbf{Y}}^l = \left(\mathbf{y}_{1,J}^{l-2}, \mathbf{y}_{2,1}^{l-2}, \mathbf{y}_{2,2}^{l-2}, \cdots, \mathbf{y}_{T,J}^{l-2}, \mathbf{y}_{1,1}^{l-1}, \mathbf{y}_{1,2}^{l-1}, \cdots, \mathbf{y}_{T,J}^{l-1}, \mathbf{y}_{1,1}^l, \mathbf{y}_{1,2}^l, \cdots, \mathbf{y}_{T,J}^l \right),$$

$$\tilde{\mathbf{N}}^l = \left(\mathbf{n}_{1,J}^{l-2}, \mathbf{n}_{2,1}^{l-2}, \mathbf{n}_{2,2}^{l-2}, \cdots, \mathbf{n}_{T,J}^{l-2}, \mathbf{n}_{1,1}^{l-1}, \mathbf{n}_{1,2}^{l-1}, \cdots, \mathbf{n}_{T,J}^{l-1}, \mathbf{n}_{1,1}^l, \mathbf{n}_{1,2}^l, \cdots, \mathbf{n}_{T,J}^l \right),$$

$$\mathbf{U}_j^l = (\mathbf{I}_N, \mathbf{P}_j^{l-1}, \mathbf{P}_j^{l-1} \mathbf{P}_j^l), j = 1, \cdots, J.$$

Proof: The result follows from the input-output relationship for any time block $l > 0$ in (17) and (18) and using simple algebra. ■

Again, notice that $\tilde{\mathbf{Y}}^l$ starts from $\mathbf{y}_{1,J}^{l-2}$ instead of $\mathbf{y}_{1,1}^{l-2}$. Other previously received signals could be considered to improve performance, but that would cause additional inter-block interference from previously transmitted signals and would increase decoding complexity. It is easy to see from Proposition 4.2 that when conditioned on the data matrices $\mathbf{P}_1^{l-1}, \mathbf{P}_1^l, \dots, \mathbf{P}_J^{l-1}, \mathbf{P}_J^l$, the matrix $\tilde{\mathbf{Y}}^l$ is Gaussian with conditional pdf

$$P\left(\tilde{\mathbf{Y}}^l \mid \mathbf{P}_1^{l-1}, \mathbf{P}_1^l, \dots, \mathbf{P}_J^{l-1}, \mathbf{P}_J^l\right) \propto \frac{\exp\left\{-\text{Tr}\left[\tilde{\mathbf{Y}}^l \cdot (\tilde{\mathbf{V}}^l)^{-1} \cdot (\tilde{\mathbf{Y}}^l)^\dagger\right]\right\}}{\left[\det(\tilde{\mathbf{V}}^l)\right]^M} \quad (35)$$

where $\tilde{\mathbf{V}}^l$ is the covariance matrix given by $\tilde{\mathbf{V}}^l = \sum_{j=1}^J (\mathbf{U}_j^l \tilde{\mathbf{A}}_j)^\dagger \cdot (\mathbf{U}_j^l \tilde{\mathbf{A}}_j) + (\text{SNR})^{-1} T_s \cdot \tilde{\mathbf{D}}$ and $\tilde{\mathbf{D}} = \text{diag}(\tau_{(1)(0)}^{-1}, \tau_{(2)(1)}^{-1}, \dots, \tau_{(3TJ-J+1)(3TJ-J)}^{-1})$ is a $3TJ - J + 1 \times 3TJ - J + 1$ diagonal matrix. Based on (35), we can define the Maximum Multiple Partition Likelihood (MMPL) decoding using Method 0 as

$$\left\{\hat{\mathbf{P}}_1^{l-1}, \hat{\mathbf{P}}_1^l, \dots, \hat{\mathbf{P}}_J^{l-1}, \hat{\mathbf{P}}_J^l\right\} = \underset{\mathbf{P}_1^{l-1}, \mathbf{P}_1^l, \dots, \mathbf{P}_J^{l-1}, \mathbf{P}_J^l}{\text{argmin}} \left\{M \cdot \ln \left[\det(\tilde{\mathbf{V}}^l)\right] + \text{Tr} \left[\tilde{\mathbf{Y}}^l \cdot (\tilde{\mathbf{V}}^l)^{-1} \cdot (\tilde{\mathbf{Y}}^l)^\dagger\right]\right\}. \quad (36)$$

The cost function of the MMPL decoder using Method 0 is a function of $\mathbf{P}_1^{l-1}, \mathbf{P}_1^l, \dots, \mathbf{P}_J^{l-1}, \mathbf{P}_J^l$, whereas the cost function of the ITIC decoder for User $j = 1, \dots, J$ using Method 0 is only a function of $\mathbf{P}_j^{l-1}, \mathbf{P}_j^l$. We can use the DP procedures in Methods 1 and 2 with the cost function of the MMPL decoder in (36) just as with the cost function of the ITIC decoder in (23). However, we need to compute and store a function of $\mathbf{P}_1^l, \dots, \mathbf{P}_J^l$ instead of $\Phi_j^l(\mathbf{P}_j^l)$ defined in (27). Similarly, Method 3 can be applied to the cost function of the MMPL decoder in (36) by using the decoded matrix for $\mathbf{P}_1^{l-1}, \dots, \mathbf{P}_J^{l-1}$ at block $l-1$ in (36) to decode $\mathbf{P}_1^l, \dots, \mathbf{P}_J^l$ at block l . The three algorithms can therefore be changed accordingly. The block diagram of the proposed differential decoders is shown in Fig. 3.

The corresponding coherent decoders for the ITIC and MMPL decoders can be derived using similar procedures to the ones described above as well. Due to space limitations, we do not

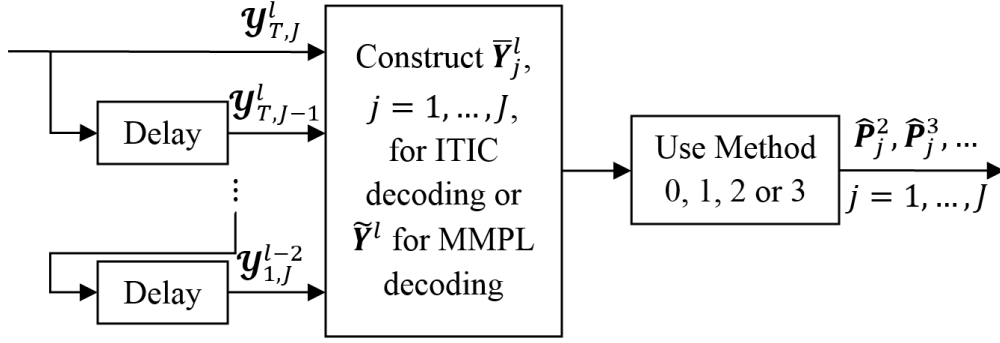


Fig. 3. Block diagram of differential decoders.

provide the details of the coherent ITIC and MMPL decoders.

V. PROOF OF FULL DIVERSITY

Diversity is defined as

$$d \triangleq - \lim_{\text{SNR} \rightarrow \infty} \frac{\log P_{err}(\text{SNR})}{\log \text{SNR}} \quad (37)$$

where $P_{err}(\text{SNR})$ represents the probability of error at the corresponding SNR. In this section, we analyze the diversity of the proposed schemes in Section IV and prove that they achieve a diversity order of MN (full diversity).

Theorem 5.1: The proposed ITIC and MMPL decoders using Method 0 achieve full diversity.

Proof: In the ITIC decoder using Method 0, we used the relationship in Equation (15) and performed noncoherent ML detection. In (15), $\mathbf{H}_j \mathbf{S}_j^{l-2}$, $\mathbf{U}_j^l \bar{\mathbf{A}}$, and $\bar{\mathbf{N}}_j^l$ can be considered as the equivalent channel, signal, and noise terms, respectively. Note that the entries of $\mathbf{H}_j \mathbf{S}_j^{l-2}$ and $\bar{\mathbf{N}}_j^l$ are samples of independent zero-mean complex Gaussian random variables. With a small abuse of the notation, let $\mathbf{U}_{j,1}^l = (\mathbf{I}_N, \mathbf{P}_{j,1}^{l-1}, \mathbf{P}_{j,1}^{l-1} \mathbf{P}_{j,1}^l)$, $\mathbf{U}_{j,2}^l = (\mathbf{I}_N, \mathbf{P}_{j,2}^{l-1}, \mathbf{P}_{j,2}^{l-1} \mathbf{P}_{j,2}^l)$ for some arbitrary data matrices $\mathbf{P}_{j,1}^{l-1}, \mathbf{P}_{j,1}^l, \mathbf{P}_{j,2}^{l-1}, \mathbf{P}_{j,2}^l$ such that $\mathbf{U}_{j,1}^l \neq \mathbf{U}_{j,2}^l$. Then, in order to prove that the ITIC decoder using Method 0 achieves a diversity order of MN , by Proposition 4 of [18],

it suffices to show that for any $U_{j,1}^l \neq U_{j,2}^l$, the following has full row rank²:

$$\begin{pmatrix} U_{j,1}^l \cdot \bar{A} \\ U_{j,2}^l \cdot \bar{A} \end{pmatrix} = \begin{pmatrix} I_N & P_{j,1}^{l-1} & P_{j,1}^{l-1} P_{j,1}^l \\ I_N & P_{j,2}^{l-1} & P_{j,2}^{l-1} P_{j,2}^l \end{pmatrix} \cdot \bar{A} = G(P_{j,1}^{l-1}, P_{j,1}^l, P_{j,2}^{l-1}, P_{j,2}^l). \quad (38)$$

By our assumption, $G(P_{j,1}^{l-1}, P_{j,1}^l, P_{j,2}^{l-1}, P_{j,2}^l)$ has full row rank when $(P_{j,1}^{l-1}, P_{j,1}^l) \neq (P_{j,2}^{l-1}, P_{j,2}^l)$ (or equivalently, $U_{j,1}^l \neq U_{j,2}^l$). Thus, the ITIC decoder using Method 0 provides full diversity. Now, note that the MMPL decoder using Method 0 is optimal among the decoders using the same set of (or a subset of) the time partitions it uses. Since the ITIC decoder using Method 0 uses a subset of the time partitions the MMPL decoder using Method 0 uses, the MMPL decoder using Method 0 must perform at least as good as the ITIC decoder using Method 0. Thus, the MMPL decoder using Method 0 must achieve full diversity as well. ■

The following theorem extends the result of Theorem 5.1 to all the proposed methods:

Theorem 5.2: If one of the proposed differential schemes using Method 0 provides full diversity, then the corresponding differential schemes using Methods 1, 2 and 3 will provide full diversity as well.

Proof: The proof is very similar to that of Theorem 5.1 in [17]. ■

Therefore, by Theorems 5.1 and 5.2, all the proposed differential schemes (i.e., ITIC and MMPL decoders using Methods 0, 1, 2 and 3) provide full diversity.

As mentioned above, in order to guarantee full diversity, we need to make sure that $G(P_1, P_2, P_3, P_4)$ has full row rank for any possible data matrices P_1, P_2, P_3, P_4 with $(P_1, P_2) \neq (P_3, P_4)$. In the following theorem we derive an equivalent condition, which can be easily verified using simple matrix operations:

Theorem 5.3: $G(P_1, P_2, P_3, P_4)$ has full row rank for any possible data matrices P_1, P_2, P_3, P_4

²The channel model used in [18] is the transposed version of ours. We have modified their results based on our channel model. We have also used the fact that $\text{rank}(X^\dagger X) = \text{rank}(X)$ for any matrix X with complex elements.

with $(\mathbf{P}_1, \mathbf{P}_2) \neq (\mathbf{P}_3, \mathbf{P}_4)$ if and only if

$$\mathbf{w} \cdot \left[\tilde{\mathbf{P}}_1 \tilde{\mathbf{P}}_2 + N \left(\mathbf{I}_N - \tilde{\mathbf{P}}_1 \right) \cdot \left(\frac{(\tilde{\mathbf{P}}_3 - \tilde{\mathbf{P}}_1)^\dagger}{\|\tilde{\mathbf{P}}_3 - \tilde{\mathbf{P}}_1\|_F^2} \right) \cdot (\tilde{\mathbf{P}}_3 \tilde{\mathbf{P}}_4 - \tilde{\mathbf{P}}_1 \tilde{\mathbf{P}}_2) \right] \neq \mathbf{w} \quad (39)$$

for any possible data matrices $\tilde{\mathbf{P}}_1, \tilde{\mathbf{P}}_2, \tilde{\mathbf{P}}_3, \tilde{\mathbf{P}}_4$ with $\tilde{\mathbf{P}}_1 \neq \tilde{\mathbf{P}}_3$, where $\mathbf{w} = \overbrace{(1, 1, \dots, 1)}^{N \text{ times}}$.

Proof: See Appendix A. ■

For instance, consider the case when the Alamouti code is used to construct the data matrices \mathbf{P}_j^l . Then, one can use Theorem 5.3 to verify that when the BPSK constellation $\{\frac{e^{j(\frac{\pi}{2})}}{\sqrt{2}}, \frac{e^{-j(\frac{\pi}{2})}}{\sqrt{2}}\}$ or the QPSK constellation $\{\frac{e^{j(\frac{\pi}{4})}}{\sqrt{2}}, \frac{e^{j(\frac{3\pi}{4})}}{\sqrt{2}}, \frac{e^{-j(\frac{3\pi}{4})}}{\sqrt{2}}, \frac{e^{-j(\frac{\pi}{4})}}{\sqrt{2}}\}$ is used, $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$ will have full row rank for any possible data matrices $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4$ with $(\mathbf{P}_1, \mathbf{P}_2) \neq (\mathbf{P}_3, \mathbf{P}_4)$. As another example, consider the case when the following 4×4 rate-one STBC is used to construct the data matrices:

$$\mathbf{P}_j^l = \begin{pmatrix} p_{j,1}^l & -p_{j,2}^l & -p_{j,3}^l & -p_{j,4}^l \\ p_{j,2}^l & p_{j,1}^l & p_{j,4}^l & -p_{j,3}^l \\ p_{j,3}^l & -p_{j,4}^l & p_{j,1}^l & p_{j,2}^l \\ p_{j,4}^l & p_{j,3}^l & -p_{j,2}^l & p_{j,1}^l \end{pmatrix}. \quad (40)$$

Note that for the constellation $\{\frac{e^{j(\frac{\pi}{2})}}{2}, \frac{e^{-j(\frac{\pi}{2})}}{2}\}$ the above STBC is orthogonal. Again, one may use Theorem 5.3 to verify that when the constellation $\{\frac{e^{j(\frac{\pi}{2})}}{2}, \frac{e^{-j(\frac{\pi}{2})}}{2}\}$ is used, $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$ will have full row rank for any possible data matrices $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4$ with $(\mathbf{P}_1, \mathbf{P}_2) \neq (\mathbf{P}_3, \mathbf{P}_4)$.

Notice that another condition for the MMPL decoders to achieve full diversity can be found directly from (33). The condition based on (33) is weaker than the condition in Theorem 5.3. However, since it depends on the data matrices of all users, it is more complicated to verify. In the coherent case, Proposition 3 of [18] can be used to obtain conditions for the coherent ITIC and MMPL decoders to achieve full diversity.

VI. SIMULATION RESULTS

In this section, we provide simulation results for the performance of the proposed differential modulation schemes using the ITIC and MMPL decoders based on Methods 1, 2 and 3. We compare the performance of our schemes to the IUIF and M3BL differential schemes presented in [17] and the synchronous coherent schemes using Zero-Forcing (ZF) and Maximum Likelihood (ML) decoding. When using Method 2 for decoding, we decode all the signals within each frame after receiving the last signal in that frame. In our simulations, the channel is quasi-static flat Rayleigh fading where the fading is constant within one frame and varies independently from one frame to another. Depending on the number of transmit antennas, we use either the Alamouti code or the 4×4 OSTBC in (40) for all users to encode and transmit 64 data matrices per user in each frame. Also, we use the BPSK and QPSK constellations described in Section V as the signal constellations for the simulations of our differential schemes at transmission rates 1 b/(s Hz) and 2 b/(s Hz), respectively. In Figs. 4-9, we consider the relative time delays between the received signals of consecutive users to be equal (i.e., $\tau_{j+1} - \tau_j = T_s/J, \forall j$). We study the effect of other relative time delays on performance in Fig. 10. In each figure, the curves for all users are identical.

Figs. 4 and 5 show BER as a function of SNR at transmission rates 1 b/(s Hz) and 2 b/(s Hz), respectively, for 2 users each equipped with 2 transmit antennas and a receiver with 2 receive antennas. In Figs. 6 and 7, we present similar results for 3 receive antennas. In Fig. 8, we provide simulation results at a transmission rate of 1 b/(s Hz) for 2 users each equipped with 4 transmit antennas and a receiver with 1 receive antenna. Note that all our schemes work for any number of receive antennas, while the low complexity differential schemes in [17] require at least J receive antennas. All simulation results demonstrate that all the proposed schemes achieve full diversity like the corresponding coherent schemes using ML decoding. On the other hand, the low complexity differential schemes in [17] only provide full transmit diversity. Additionally,

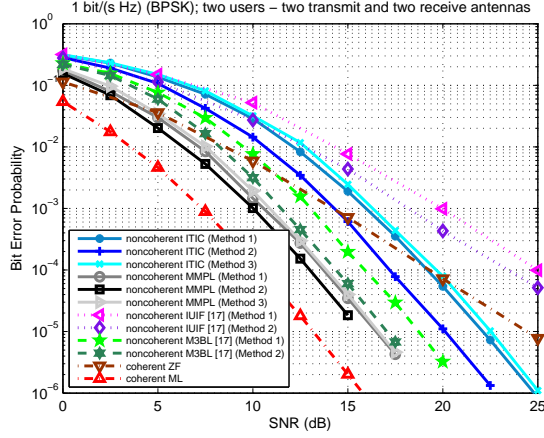


Fig. 4. Performance of coherent and differential schemes at a rate of 1 b/(s Hz) for 2 users each with 2 transmit antennas and 1 receiver with 2 receive antennas.

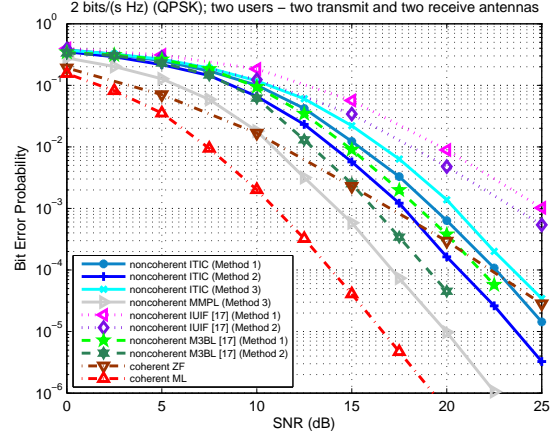


Fig. 5. Performance of coherent and differential schemes at a rate of 2 b/(s Hz) for 2 users each with 2 transmit antennas and 1 receiver with 2 receive antennas.

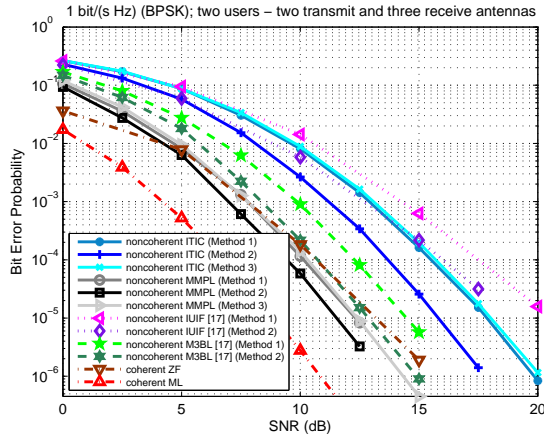


Fig. 6. Performance of coherent and differential schemes at a rate of 1 b/(s Hz) for 2 users each with 2 transmit antennas and 1 receiver with 3 receive antennas.

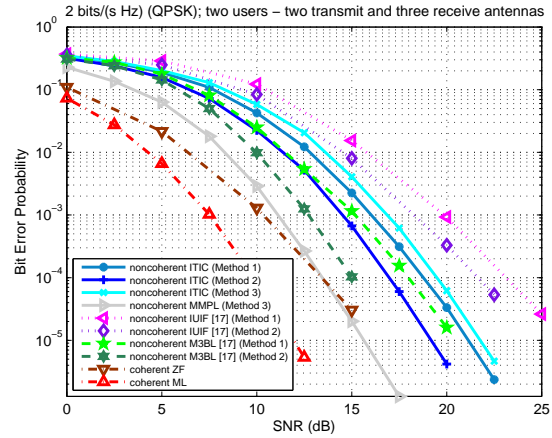


Fig. 7. Performance of coherent and differential schemes at a rate of 2 b/(s Hz) for 2 users each with 2 transmit antennas and 1 receiver with 3 receive antennas.

compared to the differential schemes in [17], the MMPL decoding schemes provide significant performance improvement. Therefore, the proposed schemes provide the possibility of a tradeoff between decoding complexity and the coding gain.

In Fig. 9, we show BER as a function of SNR at a transmission rate of 1 b/(s Hz) for 3 users each equipped with 2 transmit antennas and a receiver with 2 receive antennas. With the

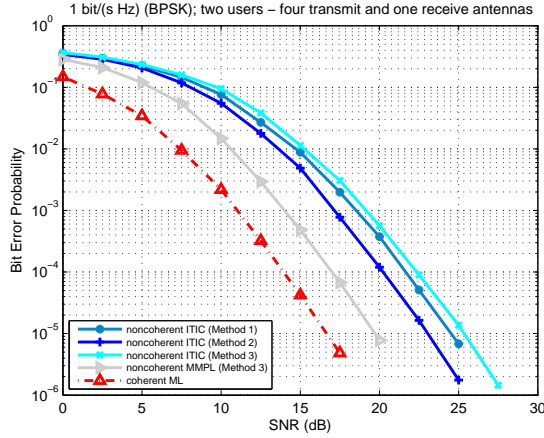


Fig. 8. Performance of coherent and differential schemes at a rate of 1 b/(s Hz) for 2 users each with 4 transmit antennas and 1 receiver with 1 receive antenna.

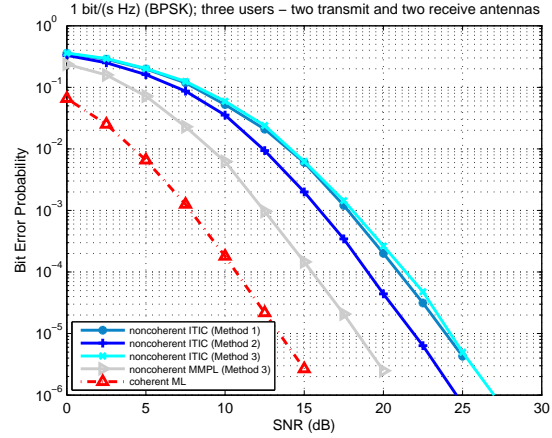


Fig. 9. Performance of coherent and differential schemes at a rate of 1 b/(s Hz) for 3 users each with 2 transmit antennas and 1 receiver with 2 receive antennas.

assumption of equal relative time delays, it can be seen from Proposition 4.1 and the covariance matrices for the noise vectors given in Section II that the effect of changing the number of users from J_1 to J_2 on the performance of the ITIC decoders is the same as that of multiplying the SNR by J_1/J_2 . This corresponds to a change of $10 \log_{10}(J_1/J_2)$ dB in performance. As expected, the performances of the ITIC decoders in Fig. 4 for 2 users are $10 \log_{10}(3/2) \approx 1.8$ dB better than those of Fig. 9 for 3 users. All simulations show that the effect of error propagation on the performance of the proposed schemes using Method 3 is very small. Our schemes using Method 3 have lower decoding complexity compared to their corresponding schemes using Method 1, yet the proposed schemes using Method 3 provide almost the same performance as their corresponding schemes using Method 1.

Finally, we compare the performance of our differential schemes with different relative time delays between the received signals. Again, we consider a system with 2 users each equipped with 2 transmit antennas and a receiver with 2 receive antennas. Fig. 10 shows the performance of the ITIC and MMPL decoders using Method 3 for different values of $\Delta\tau = \tau_2 - \tau_1$ at a transmission rate of 1 b/(s Hz). The results for our decoding schemes using Methods 0, 1 and

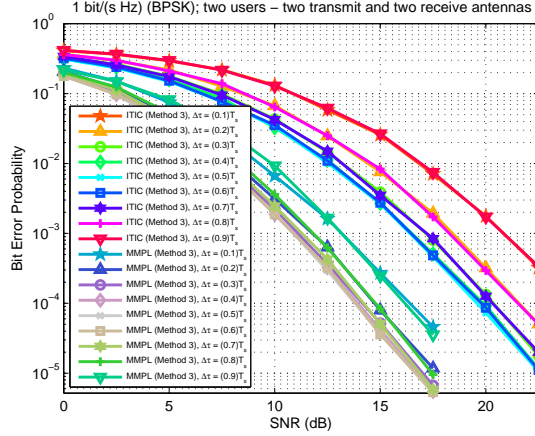


Fig. 10. Comparison of our schemes using Method 3 for different relative time delays $\Delta\tau = \tau_2 - \tau_1$ at a rate of 1 b/(s Hz) for 2 users each with 2 transmit antennas and 1 receiver with 2 receive antennas.

2 are similar. It is evident from the simulations that the proposed schemes perform best when $\Delta\tau = T_s/2$, that is, when the signals of the two users are received with a time difference of half a symbol. Moreover, for values of $\Delta\tau$ close to $T_s/2$, the performance of our schemes is close to the best performance for $\Delta\tau = T_s/2$ and deviates from the best performance more quickly as $\Delta\tau$ deviates from $T_s/2$. This is in line with capacity results reported in [19] where $\Delta\tau = T_s/2$ provides the highest value of channel capacity in a two-user MAC.

VII. CONCLUSIONS

We introduced differential detection schemes for asynchronous multi-user MIMO systems based on orthogonal STBCs where neither the transmitters nor the receiver knows the CSI. We first presented schemes with simple differential encoding and low complexity differential decoding algorithms by performing interference cancelation in time and employing different decoding methods. The decoding complexity of these schemes increases linearly with the number of users. We then presented additional differential decoding schemes that achieve significantly higher coding gains compared to our low complexity schemes. Simulation results show that they

also outperform the existing synchronous differential schemes. The proposed schemes work for any square OSTBC, any number of users, and a receiver with any number of receive antennas. Furthermore, we proved analytically that our schemes achieve full diversity with the appropriate choice of constellations. To the best of our knowledge, the proposed differential modulation schemes are the first differential schemes for asynchronous multi-user communication systems.

APPENDIX A

PROOF OF THEOREM 5.3

We need the following property to prove the theorem:

Lemma A.1: Let $\mathbf{X}_1, \mathbf{X}_2$ be distinct $N \times N$ matrices such that $(\mathbf{X}_2 - \mathbf{X}_1)^\dagger \cdot (\mathbf{X}_2 - \mathbf{X}_1) = \frac{\|\mathbf{X}_2 - \mathbf{X}_1\|_F^2}{N} \cdot \mathbf{I}_N$. Then,

$$\begin{pmatrix} \mathbf{I}_N & \mathbf{X}_1 \\ \mathbf{I}_N & \mathbf{X}_2 \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{I}_N + \mathbf{X}_1 \bar{\mathbf{X}} & -\mathbf{X}_1 \bar{\mathbf{X}} \\ -\bar{\mathbf{X}} & \bar{\mathbf{X}} \end{pmatrix} \quad (41)$$

where $\bar{\mathbf{X}} = \frac{N(\mathbf{X}_2 - \mathbf{X}_1)^\dagger}{\|\mathbf{X}_2 - \mathbf{X}_1\|_F^2}$.

Proof: The result can be easily proven by showing that

$$\begin{pmatrix} \mathbf{I}_N + \mathbf{X}_1 \bar{\mathbf{X}} & -\mathbf{X}_1 \bar{\mathbf{X}} \\ -\bar{\mathbf{X}} & \bar{\mathbf{X}} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{I}_N & \mathbf{X}_1 \\ \mathbf{I}_N & \mathbf{X}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_N & \mathbf{0}_N \\ \mathbf{0}_N & \mathbf{I}_N \end{pmatrix} = \mathbf{I}_{2N}. \quad (42)$$

■

To prove Theorem 5.3, we consider two cases:

Case 1: We first consider the case when $\mathbf{P}_1 \neq \mathbf{P}_3$. Since $(\mathbf{P}_3 - \mathbf{P}_1)^\dagger \cdot (\mathbf{P}_3 - \mathbf{P}_1) = \frac{\|\mathbf{P}_3 - \mathbf{P}_1\|_F^2}{N} \cdot \mathbf{I}_N$, by Lemma A.1, $\begin{pmatrix} \mathbf{I}_N & \mathbf{P}_1 \\ \mathbf{I}_N & \mathbf{P}_3 \end{pmatrix}$ is invertible. Also, since its inverse must be a full rank matrix, multiplying its inverse by $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$ must result in a matrix with the same rank as $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$. Therefore, using Lemma A.1 and the definition of $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$ in (14),

by multiplying $G(P_1, P_2, P_3, P_4)$ by $\begin{pmatrix} I_N & P_1 \\ I_N & P_3 \end{pmatrix}^{-1}$ from the left we obtain

$$\begin{aligned} \begin{pmatrix} I_N & P_1 \\ I_N & P_3 \end{pmatrix}^{-1} \cdot G(P_1, P_2, P_3, P_4) &= \begin{pmatrix} I_N + P_1 \left(\frac{N(P_3 - P_1)^\dagger}{\|P_3 - P_1\|_F^2} \right) & -P_1 \left(\frac{N(P_3 - P_1)^\dagger}{\|P_3 - P_1\|_F^2} \right) \\ -\frac{N(P_3 - P_1)^\dagger}{\|P_3 - P_1\|_F^2} & \frac{N(P_3 - P_1)^\dagger}{\|P_3 - P_1\|_F^2} \end{pmatrix} \cdot \begin{pmatrix} I_N & P_1 & P_1 P_2 \\ I_N & P_3 & P_3 P_4 \end{pmatrix} \cdot \bar{A} \\ &= \begin{pmatrix} I_N & 0_N & P_1 P_2 - N P_1 \cdot \left(\frac{(P_3 - P_1)^\dagger}{\|P_3 - P_1\|_F^2} \right) \cdot (P_3 P_4 - P_1 P_2) \\ 0_N & I_N & N \left(\frac{(P_3 - P_1)^\dagger}{\|P_3 - P_1\|_F^2} \right) \cdot (P_3 P_4 - P_1 P_2) \end{pmatrix} \cdot \bar{A}, \end{aligned} \quad (43)$$

which must be of the same rank as $G(P_1, P_2, P_3, P_4)$. Now, Let B_1 and B_1^{-1} be $3N - 1 \times 3N - 1$ matrices given by

$$B_1 = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & -1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \quad B_1^{-1} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 0 & 1 & 1 & \cdots & 1 & 1 \\ 0 & 0 & 1 & \cdots & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 1 \\ 0 & 0 & 0 & \cdots & 1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}. \quad (44)$$

Note that B_1^{-1} is the inverse of B_1 . Again, since B_1^{-1} is a full rank matrix, multiplying it by (43) will result in a matrix with the same rank as (43). Therefore, multiplying (43) by B_1^{-1} from the right yields a matrix with the same rank as $G(P_1, P_2, P_3, P_4)$, given by

$$\begin{pmatrix} I_N & P_1 \\ I_N & P_3 \end{pmatrix}^{-1} \cdot G(P_1, P_2, P_3, P_4) \cdot B_1^{-1} = \begin{pmatrix} I_N & 0_N & P_1 P_2 - N P_1 \cdot \left(\frac{(P_3 - P_1)^\dagger}{\|P_3 - P_1\|_F^2} \right) \cdot (P_3 P_4 - P_1 P_2) \\ 0_N & I_N & N \left(\frac{(P_3 - P_1)^\dagger}{\|P_3 - P_1\|_F^2} \right) \cdot (P_3 P_4 - P_1 P_2) \end{pmatrix} \cdot B_2 \quad (45)$$

where B_2 is the $3N \times 3N - 1$ matrix

$$B_2 = \bar{A} \cdot B_1^{-1} = \begin{pmatrix} -1 & -1 & -1 & \cdots & -1 & -1 & -1 \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}. \quad (46)$$

Now, consider the RHS of (45) and let

$$\begin{pmatrix} \mathbf{P}_1 \mathbf{P}_2 - N \mathbf{P}_1 \cdot \left(\frac{(\mathbf{P}_3 - \mathbf{P}_1)^\dagger}{\|\mathbf{P}_3 - \mathbf{P}_1\|_F^2} \right) \cdot (\mathbf{P}_3 \mathbf{P}_4 - \mathbf{P}_1 \mathbf{P}_2) \\ N \left(\frac{(\mathbf{P}_3 - \mathbf{P}_1)^\dagger}{\|\mathbf{P}_3 - \mathbf{P}_1\|_F^2} \right) \cdot (\mathbf{P}_3 \mathbf{P}_4 - \mathbf{P}_1 \mathbf{P}_2) \end{pmatrix} = \begin{pmatrix} \beta_{1,1} & \beta_{1,2} & \beta_{1,3} & \cdots & \beta_{1,N} \\ \beta_{2,1} & \beta_{2,2} & \beta_{2,3} & \cdots & \beta_{2,N} \\ \beta_{3,1} & \beta_{3,2} & \beta_{3,3} & \cdots & \beta_{3,N} \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \beta_{2N,1} & \beta_{2N,2} & \beta_{2N,3} & \cdots & \beta_{2N,N} \end{pmatrix}. \quad (47)$$

By plugging (47) into (45) and using simple algebra, we can write (45) as

$$\begin{pmatrix} \mathbf{I}_N & \mathbf{P}_1 \\ \mathbf{I}_N & \mathbf{P}_3 \end{pmatrix}^{-1} \cdot \underbrace{\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)}_{2N-1 \text{ times}} \cdot \mathbf{B}_1^{-1} = \begin{pmatrix} -1 & -1 & -1 & \cdots & -1 & \beta_{1,1} - 1 & \beta_{1,2} - 1 & \beta_{1,3} - 1 & \cdots & \beta_{1,N} - 1 \\ 1 & 0 & 0 & \cdots & 0 & \beta_{2,1} & \beta_{2,2} & \beta_{2,3} & \cdots & \beta_{2,N} \\ 0 & 1 & 0 & \cdots & 0 & \beta_{3,1} & \beta_{3,2} & \beta_{3,3} & \cdots & \beta_{3,N} \\ 0 & 0 & 1 & \cdots & 0 & \beta_{4,1} & \beta_{4,2} & \beta_{4,3} & \cdots & \beta_{4,N} \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & \cdots & 1 & \beta_{2N,1} & \beta_{2N,2} & \beta_{2N,3} & \cdots & \beta_{2N,N} \end{pmatrix}. \quad (48)$$

Let \mathbf{r}_i , $i = 1, \dots, 2N$, denote the i th row of (48). Then, the linear combination of $\mathbf{r}_1, \dots, \mathbf{r}_{2N}$ with coefficients $\lambda_1, \lambda_2, \dots, \lambda_{2N}$, which are not all zero, is given by

$$\mathbf{r} = \sum_{i=1}^{2N} \lambda_i \mathbf{r}_i = \left(\lambda_2 - \lambda_1, \dots, \lambda_{2N} - \lambda_1, -\lambda_1 + \sum_{i=1}^{2N} \lambda_i \beta_{i,1}, \dots, -\lambda_1 + \sum_{i=1}^{2N} \lambda_i \beta_{i,N} \right). \quad (49)$$

Note that \mathbf{r} is equal to the zero vector if and only if $\lambda_1 = \lambda_2 = \dots = \lambda_{2N}$ and $\sum_{i=1}^{2N} \beta_{i,1} = \sum_{i=1}^{2N} \beta_{i,2} = \dots = \sum_{i=1}^{2N} \beta_{i,N} = 1$. This means that the rows of (48) are linearly dependent if and only if $\sum_{i=1}^{2N} \beta_{i,1} = \sum_{i=1}^{2N} \beta_{i,2} = \dots = \sum_{i=1}^{2N} \beta_{i,N} = 1$. Using (47), this implies that (48), and thus $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$, has full row rank if and only if

$$\underbrace{(1, 1, \dots, 1)}_{2N \text{ times}} \cdot \begin{pmatrix} \mathbf{P}_1 \mathbf{P}_2 - N \mathbf{P}_1 \cdot \left(\frac{(\mathbf{P}_3 - \mathbf{P}_1)^\dagger}{\|\mathbf{P}_3 - \mathbf{P}_1\|_F^2} \right) \cdot (\mathbf{P}_3 \mathbf{P}_4 - \mathbf{P}_1 \mathbf{P}_2) \\ N \left(\frac{(\mathbf{P}_3 - \mathbf{P}_1)^\dagger}{\|\mathbf{P}_3 - \mathbf{P}_1\|_F^2} \right) \cdot (\mathbf{P}_3 \mathbf{P}_4 - \mathbf{P}_1 \mathbf{P}_2) \end{pmatrix} \neq \underbrace{(1, 1, \dots, 1)}_{N \text{ times}}. \quad (50)$$

Then, it is easy to see that (50) holds, and thus $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$ has full row rank, for any possible data matrices $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4$ with $\mathbf{P}_1 \neq \mathbf{P}_3$ if and only if (39) holds for any possible

data matrices $\tilde{\mathbf{P}}_1, \tilde{\mathbf{P}}_2, \tilde{\mathbf{P}}_3, \tilde{\mathbf{P}}_4$ with $\tilde{\mathbf{P}}_1 \neq \tilde{\mathbf{P}}_3$. This means that (39) is a necessary and sufficient condition for $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$ to have full row rank in Case 1.

Case 2: We now consider the case when $\mathbf{P}_1 = \mathbf{P}_3$. Since $(\mathbf{P}_1, \mathbf{P}_2) \neq (\mathbf{P}_3, \mathbf{P}_4)$, this implies that $\mathbf{P}_2 \neq \mathbf{P}_4$. Also, since $[\mathbf{P}_1(\mathbf{P}_4 - \mathbf{P}_2)]^\dagger \cdot [\mathbf{P}_1(\mathbf{P}_4 - \mathbf{P}_2)] = \frac{\|\mathbf{P}_1(\mathbf{P}_4 - \mathbf{P}_2)\|_F^2}{N} \cdot \mathbf{I}_N$, by Lemma A.1, $\begin{pmatrix} \mathbf{I}_N & \mathbf{P}_1\mathbf{P}_2 \\ \mathbf{I}_N & \mathbf{P}_1\mathbf{P}_4 \end{pmatrix}$ is invertible. Again, since its inverse must be a full rank matrix, multiplying its inverse by $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$ must result in a matrix with the same rank as $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$. Therefore, by multiplying $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$ by $\begin{pmatrix} \mathbf{I}_N & \mathbf{P}_1\mathbf{P}_2 \\ \mathbf{I}_N & \mathbf{P}_1\mathbf{P}_4 \end{pmatrix}^{-1}$ from the left we obtain

$$\begin{aligned} & \begin{pmatrix} \mathbf{I}_N & \mathbf{P}_1\mathbf{P}_2 \\ \mathbf{I}_N & \mathbf{P}_1\mathbf{P}_4 \end{pmatrix}^{-1} \cdot \mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4) \\ &= \begin{pmatrix} \mathbf{I}_N + \mathbf{P}_1\mathbf{P}_2 \left(\frac{N[\mathbf{P}_1(\mathbf{P}_4 - \mathbf{P}_2)]^\dagger}{\|\mathbf{P}_1(\mathbf{P}_4 - \mathbf{P}_2)\|_F^2} \right) & -\mathbf{P}_1\mathbf{P}_2 \left(\frac{N[\mathbf{P}_1(\mathbf{P}_4 - \mathbf{P}_2)]^\dagger}{\|\mathbf{P}_1(\mathbf{P}_4 - \mathbf{P}_2)\|_F^2} \right) \\ -\frac{N[\mathbf{P}_1(\mathbf{P}_4 - \mathbf{P}_2)]^\dagger}{\|\mathbf{P}_1(\mathbf{P}_4 - \mathbf{P}_2)\|_F^2} & \frac{N[\mathbf{P}_1(\mathbf{P}_4 - \mathbf{P}_2)]^\dagger}{\|\mathbf{P}_1(\mathbf{P}_4 - \mathbf{P}_2)\|_F^2} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{I}_N & \mathbf{P}_1 & \mathbf{P}_1\mathbf{P}_2 \\ \mathbf{I}_N & \mathbf{P}_1 & \mathbf{P}_1\mathbf{P}_4 \end{pmatrix} \cdot \bar{\mathbf{A}} \\ &= \begin{pmatrix} \mathbf{I}_N & \mathbf{P}_1 & \mathbf{0}_N \\ \mathbf{0}_N & \mathbf{0}_N & \mathbf{I}_N \end{pmatrix} \cdot \bar{\mathbf{A}}, \end{aligned} \tag{51}$$

which must be of the same rank as $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$. Once again, since \mathbf{B}_1^{-1} is a full rank matrix, multiplying (51) by \mathbf{B}_1^{-1} from the right yields a matrix with the same rank as (51), and thus $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$, given by

$$\begin{pmatrix} \mathbf{I}_N & \mathbf{P}_1\mathbf{P}_2 \\ \mathbf{I}_N & \mathbf{P}_1\mathbf{P}_4 \end{pmatrix}^{-1} \cdot \mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4) \cdot \mathbf{B}_1^{-1} = \begin{pmatrix} \mathbf{I}_N & \mathbf{P}_1 & \mathbf{0}_N \\ \mathbf{0}_N & \mathbf{0}_N & \mathbf{I}_N \end{pmatrix} \cdot \mathbf{B}_2. \tag{52}$$

Then proceeding similarly to the procedure described in (47)-(50) for Case 1, we find that $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$ has full row rank if and only if $\mathbf{w} \cdot \mathbf{P}_1 \neq \mathbf{w}$. Note that this condition is a special case of (39) when $\tilde{\mathbf{P}}_2 = \tilde{\mathbf{P}}_4 = \mathbf{P}_1$. Therefore, (39) is a sufficient condition for $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$ to have full row rank in Case 2. Also, we showed that (39) is a necessary and sufficient condition for $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$ to have full row rank in Case 1. Thus, (39) is a necessary and sufficient condition in the general case for $\mathbf{G}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$ to have full row rank for any possible data matrices $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4$ with $(\mathbf{P}_1, \mathbf{P}_2) \neq (\mathbf{P}_3, \mathbf{P}_4)$.

REFERENCES

- [1] H. Jafarkhani, “*Space-time coding: theory and practice*,” Cambridge University Press, 2005.
- [2] S. M. Alamouti, “A simple transmit diversity technique for wireless communications,” *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451-1458, Oct. 1998.
- [3] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, “Space-time block codes from orthogonal designs,” *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456-1467, Jul. 1999.
- [4] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, “Diversity versus channel knowledge at finite block-length,” in *proc. 2012 IEEE Inf. Theory Workshop*, vol., no., pp.572,576, 3-7 Sept. 2012.
- [5] V. Tarokh and H. Jafarkhani, “A differential detection scheme for transmit diversity,” *IEEE J. Sel. Areas Commun.*, vol. 18, no. 7, pp. 1169-1174, Jul. 2000.
- [6] H. Jafarkhani and V. Tarokh, “Multiple transmit antenna differential detection from generalized orthogonal designs,” *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2626-2631, Sep. 2001.
- [7] G. Ganesan and P. Stoica, “Differential modulation using space-time block codes,” *IEEE Signal Process. Lett.*, vol. 9, no. 2, pp. 57-60, Feb. 2002.
- [8] Z. Chen, G. Zhu, D. Qu, and Y. Liu, “General differential space-time modulation,” in *Proc. 2003 IEEE Global Telecommunications Conf.*, vol. 1, pp. 282-286.
- [9] M. Tao and R. S. Cheng, “Differential space-time block codes,” in *Proc. 2001 IEEE Global Telecommunications Conf.*, vol. 2, pp. 1098-1102.
- [10] H. Jafarkhani, “A quasi-orthogonal space-time block code,” *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 14, Jan. 2001.
- [11] Y. Zhu and H. Jafarkhani, “Differential modulation based on quasi-orthogonal codes,” *IEEE Trans. Wireless Commun.*, pp. 3005-3017, Nov. 2005.
- [12] A. F. Naguib, N. Seshadri, and A. R. Calderbank, “Applications of space-time block codes and interference suppression for high capacity and high data rate wireless systems,” in *Proc. 32nd Asilomar Conf. Signals, Systems and Computers*, pp. 1803-1810, 1998.
- [13] A. Stamoulis, N. Al-Dhahir and A. R. Calderbank, “Further results on interference cancellation and space-time block codes,” in *Proc. 35th Asilomar Conf. Signals, Systems and Computers*, pp. 257-262, Oct. 2001.
- [14] J. Kazemitabar and H. Jafarkhani, “Multiuser interference cancellation and detection for users with more than two transmit antennas,” *IEEE Trans. Commun.*, vol. 56, no. 4, pp. 574-583, Apr. 2008.
- [15] M. E. Gärtner and H. Bölcskei, “Multiuser space-time/frequency code design,” in *Proc. 2006 IEEE ISIT*, pp. 2819-2823.
- [16] M. R. Bhatnagar and A. Hjørungnes, “Differential Coding for MAC Based Two-User MIMO Communication Systems,” *IEEE Trans. Wireless Commun.*, vol.11, no.1, pp.9-14, Jan. 2012.
- [17] S. Poorkasmaei and H. Jafarkhani, “Orthogonal Differential Modulation for MIMO Multiple Access Channels with Two Users,” *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2374-2384, June 2013.

- [18] M. Brehler and M.K. Varanasi, "Asymptotic error probability analysis of quadratic receivers in Rayleigh-fading channels with applications to a unified analysis of coherent and noncoherent space-time receivers," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2383-2399, Sep. 2001.
- [19] S. Verdu, "The capacity region of the symbol-asynchronous Gaussian multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 35, no. 4, pp. 733-751, Jul. 1989.